

# Morphic images of episturmian words having finite palindromic defect

Štěpán Starosta  
Czech Technical University in Prague  
Czech Republic

## Abstract

We study morphisms from certain classes and their action on episturmian words. The first class is  $P_{ret}$ . In general, a morphism of class  $P_{ret}$  can map an infinite word having zero palindromic defect to a word having infinite palindromic defect. We show that the image of an episturmian word, which has zero palindromic defect, under a morphism of class  $P_{ret}$  has always its palindromic defect finite. We also focus on letter-to-letter morphisms to binary alphabet: we show that images of ternary episturmian words under such morphisms have zero palindromic defect. These results contribute to the study of an unsolved question of characterization of morphisms that preserve finite (resp. zero) palindromic defect. They also enable us to construct new examples of binary  $H$ -rich and almost  $H$ -rich words, where  $H = \{\text{Id}, R, E, RE\}$  is the group generated by both involutory antimorphisms on a binary alphabet.

## 1 Introduction

In combinatorics on words, the most famous class of words probably is the class of Sturmian words: aperiodic words having minimal factor complexity possible (see [28]). Sturmian words are profoundly studied and many generalizations are known, see for instance [4]. One such generalization of Sturmian words are episturmian words. Episturmian words were inspired by Arnoux–Rauzy words (see [32, 1]). An infinite word over a  $k$ -letter alphabet is *episturmian* if it is closed under reversal and has at most one left special factor of each length. Refer for instance to [17, 24, 19] for more results on this class.

A notion related to the study of episturmian words is a *palindrome* – a word equal to its reversal. Episturmian words are rich in palindromes: they contain the maximum number of distinct palindromic factors possible. Precisely, we say that a finite word  $w$  is *rich* if it contains exactly  $|w| + 1$  distinct palindromic factors, which is the upper bound for the number of distinct palindromic factors in a finite word of length  $|w|$  (see [17]). The notion is extended to infinite words: an infinite word is *rich* if every its factor is rich.

In the context of this upper bound on the number of palindromic factors, a measure of the count of missing palindromic factors was introduced in [9]: the *palindromic defect*  $D(w)$  of a finite word  $w$  is

$$D(w) = |w| + 1 - \#\text{Pal}(w),$$

where  $\text{Pal}(w)$  is the set of all palindromic factors of  $w$ . The palindromic defect of an infinite word  $\mathbf{u}$  is defined by  $D(\mathbf{u}) = \sup\{D(w) : w \text{ is a factor of } \mathbf{u}\}$ . If  $D(\mathbf{u})$  is finite, we say that  $\mathbf{u}$  is *almost rich*. (If it is zero, then  $\mathbf{u}$  is rich as already mentioned.)

Besides episturmian words, examples of rich words include some well-explored word classes such as words coding symmetric interval exchange transformation (see [2]) and words coding rotation on two intervals (see [8]). Properties and characterizations of rich words are studied for instance in [20, 4, 14, 11]. Examples of words that have finite and nonzero palindromic defect are also known. Such words can be constructed using iterated palindromic closure: let  $w_0 \in \mathcal{A}^*$  with  $\mathcal{A}$  being an alphabet. Set  $w_i = (w_{i-1}\delta_i)^R$  where  $\delta_i \in \mathcal{A}$  and  $w^R$  is the shortest palindrome having  $w$  as a prefix (i.e., the *palindromic closure* of  $w$ ). The infinite word having  $w_i$  as its prefix for all  $i$  is the *standard word with seed  $w_0$  and directive sequence  $(\delta_i)_{i=0}^{+\infty}$* . It follows from [12] that such a word is almost rich. General properties and characterizations of almost rich words are studied in [20, 5, 6].

In this paper, we study richness and almost richness of images of episturmian words by a morphism from a specific class. Our first result states that we obtain an almost rich word while using a morphism of class  $P_{ret}$  introduced in [5] (see Section 2.2 later for the definition).

**Theorem 1.** *Let  $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$  be an episturmian word and  $\pi : \mathcal{A}^* \rightarrow \mathcal{B}^*$  be a morphism of class  $P_{ret}$ . The word  $\pi(\mathbf{u})$  is almost rich.*

The second main result involves a letter-to-letter projection of a ternary episturmian word to binary alphabet. We use the following definition for a such a projection.

**Definition 2.** Let  $\mathcal{A}$  be an alphabet and  $\mathcal{A}'$  its proper subset. A morphism  $\zeta : \mathcal{A} \rightarrow \{A, B\}$  defined by

$$\zeta : a \mapsto \begin{cases} A & \text{if } a \in \mathcal{A}', \\ B & \text{otherwise.} \end{cases}$$

is called a *binary projection* from  $\mathcal{A}$ .

The second main result states that we obtain a rich word by projecting a ternary episturmian word to a binary alphabet.

**Theorem 3.** *Let  $\mathbf{u}$  be an episturmian word over a ternary alphabet  $\mathcal{A}$  and  $\zeta$  be a binary projection from  $\mathcal{A}$ . The word  $\zeta(\mathbf{u})$  is rich.*

Our motivation for these results is to find new binary words which are rich in a generalized sense — with respect to both symmetries given by the involutive antimorphisms on a binary alphabet: the reversal  $R$  and the exchange of letters  $E$ . We give a definition

in Section 4, see also [29, 31] for more information on this generalization. To construct new binary words rich in this generalized sense, we use the recent results of [30] which provides theorems that relate the classical richness and the generalized richness on the binary alphabet.

Our computer experiments suggest that we can improve Theorem 3: we can drop the requirement on the size of the alphabet  $\mathcal{A}$ . We state this hypothesis in the last section.

The paper is organized as follows. The next section contains some necessary definitions and basic results. Section 3 contains overview of results on episturmian words and proofs of the main results. Finally, Section 4 contains an application of the main results: a construction of binary words which are rich and almost rich in the generalized sense. The last section states some comments and open questions.

## 2 Preliminaries

### 2.1 Notions of combinatorics on words

Let  $\mathcal{A}$  be an *alphabet* — a finite set of *letters*. A finite sequence  $w = w_0w_1 \cdots w_{n-1}$  with  $w_i \in \mathcal{A}$  for all  $i$  is a *finite word*. The *length* of the word  $w$  is denoted  $|w|$  and equals  $n$ . The unique word of length 0 is the *empty word*, it is denoted  $\varepsilon$ . The set  $\mathcal{A}^*$  is the set of all finite words over  $\mathcal{A}$ . The set  $\mathcal{A}^*$  equipped with concatenation forms a free monoid with the neutral element  $\varepsilon$ . A word  $v \in \mathcal{A}^*$  is a *factor* of a word  $w \in \mathcal{A}^*$  if  $w = uvz$  for some words  $u, z \in \mathcal{A}^*$ . If  $u = \varepsilon$ , then  $v$  is a *prefix* of  $w$ ; if  $z = \varepsilon$ , then  $v$  is a *suffix* of  $w$ . If  $w$  is of the form  $w = vz$ , then  $z$  is denoted  $z = v^{-1}w$  and the word  $v^{-1}wv$  is a *conjugate* of the word  $w$ .

An *infinite word* over  $\mathcal{A}$  is a sequence  $\mathbf{u} = (u_j)_{j \in \mathbb{N}} = u_0u_1u_2 \dots$ . The set of all infinite words over  $\mathcal{A}$  is denoted by  $\mathcal{A}^{\mathbb{N}}$ . A finite word  $w \in \mathcal{A}^*$  of length  $n$  is a *factor* of  $\mathbf{u} = (u_j)_{j \in \mathbb{N}}$  if there exists an integer  $i$  such that  $w = u_iu_{i+1} \cdots u_{i+n-1}$ . The integer  $i$  is an *occurrence* of the factor  $w$  in  $\mathbf{u}$ . The *language* of  $\mathbf{u}$  is the set of all its factors and is denoted by  $\mathcal{L}(\mathbf{u})$ . Given  $a \in \mathcal{A}$  and  $w \in \mathcal{A}^*$ , if  $wa \in \mathcal{L}(\mathbf{u})$ , then  $wa$  is a *right extension* of the factor  $w$ . The set of all right extensions of  $w$  is denoted  $\text{Rext}(w)$ . Any factor of  $\mathbf{u}$  has at least one right extension. If  $w$  has at least two right extensions, it is *right special*. The notions of *left extension* of a factor, *left special* factor and  $\text{Lext}(w)$  are defined analogously. A factor  $w$  which is left and right special is *bispecial*.

An infinite word  $\mathbf{u}$  is *recurrent* if any factor of  $\mathbf{u}$  has infinitely many occurrences in  $\mathbf{u}$ . If for every factor the sequence of all its consecutive occurrences has its first difference bounded, then the word  $\mathbf{u}$  is *uniformly recurrent*. Let  $rw$  be a factor of  $\mathbf{u}$  such that  $rw$  has a prefix  $w$  and  $w$  occurs as a factor in  $rw$  exactly twice. Such a word  $r$  is a *return word* of  $w$  and the word  $rw$  is a *complete return word* of  $w$ . An infinite word  $\mathbf{u}$  is uniformly recurrent if and only if every its factor has finitely many return words.

An infinite word  $\mathbf{u}$  is *eventually periodic* if there exist words  $p$  and  $z$  such that  $\mathbf{u} = pz^{\omega} \dots = pz^{\omega}$ . It is *periodic* if  $p = \varepsilon$ . If an infinite word is not eventually periodic, it is *aperiodic*.

We denote by  $\mathcal{C}_{\mathbf{u}}$  the mapping  $\mathbb{N} \rightarrow \mathbb{N}$  which is determined by  $\mathcal{C}_{\mathbf{u}}(n) = \#(\mathcal{L}(\mathbf{u}) \cap \mathcal{A}^n)$ , i.e., it counts the factors of length  $n$  of the word  $\mathbf{u}$ . This mapping is the *factor complexity* of  $\mathbf{u}$ .

Given two alphabets  $\mathcal{A}$  and  $\mathcal{B}$ , a mapping  $\mu : \mathcal{A}^* \rightarrow \mathcal{B}^*$  is a *morphism* if  $\mu(wv) = \mu(w)\mu(v)$  for all  $w, v \in \mathcal{A}^*$ . It is an *antimorphism* if  $\mu(wv) = \mu(v)\mu(w)$  for all  $w, v \in \mathcal{A}^*$ . An infinite word  $\mathbf{u}$  is *closed under the mapping*  $\mu$  if for any factor  $w \in \mathcal{L}(\mathbf{u})$  we have also  $\mu(w) \in \mathcal{L}(\mathbf{u})$ . A morphism  $\nu : \mathcal{A}^* \rightarrow \mathcal{B}^*$  is a *conjugate morphism* to a morphism  $\mu$  if there exists a word  $w \in \mathcal{B}^*$  such that for any letter  $a \in \mathcal{A}$  we have  $w\mu(a) = \nu(a)w$ .

An antimorphism  $\Psi$  is *involutory* if  $\Psi^2 = \text{Id}$ . The most frequent involutory antimorphism is the reversal mapping  $R$  which is given by

$$R(w_0w_1 \cdots w_{n-1}) = w_{n-1} \cdots w_0 \quad \text{with } w_i \in \mathcal{A}.$$

It can be easily seen that if a word  $\mathbf{u}$  is closed under an involutory antimorphism, then  $\mathbf{u}$  is recurrent.

If  $w = R(w)$ , then  $w$  is a *palindrome*. The set of all palindromes occurring as factors of a finite word  $w$  is denoted  $\text{Pal}(w)$ . The *palindromic complexity* of an infinite word  $\mathbf{u}$  is the mapping  $P_{\mathbf{u}} : \mathbb{N} \rightarrow \mathbb{N}$  defined by  $P_{\mathbf{u}}(n) = \#\{p \in \mathcal{L}(\mathbf{u}) : p = R(p), |p| = n\}$ , i.e., the number of palindromic factors of length  $n$ .

A palindrome  $w$  is *centered at*  $x \in \mathcal{A} \cup \{\varepsilon\}$  if  $w = vxR(v)$  for some word  $v$ . Obviously, a palindrome is centered at  $\varepsilon$  if and only if it is of even length.

Let  $w \in \mathcal{A}^*$ . The quantity measuring the number of missing palindromic factors in  $w$  is the (*palindromic*) *defect* of  $w$ , introduced in [9] and given by

$$D(w) = |w| + 1 - \#\text{Pal}(w).$$

The notion of defect is extended to infinite words in the following way: let  $\mathbf{u}$  be an infinite word, we set

$$D(\mathbf{u}) = \sup\{D(w) : w \in \mathcal{L}(\mathbf{u})\}.$$

If the defect of a finite or infinite word is zero, the word is fully saturated by distinct palindromic factors and is said to be *rich* (or sometimes *full*). If its defect is finite, it is *almost rich*.

To prove (almost) richness of a word we will use the characterization of rich words given in [3]. It exploits the notion of bilateral order  $b(w)$  of a factor  $w$  and palindromic extension of a palindrome. The *bilateral order* was introduced in [15] as

$$b(w) = \#\{awb \in \mathcal{L}(\mathbf{u}) : a, b \in \mathcal{A}\} - \#\text{Rext}(w) - \#\text{Lext}(w) + 1. \quad (1)$$

If  $a$  and  $b$  are letters, we say that  $awb$  is a *both-sided extension* of  $w$  if  $awb \in \mathcal{L}(\mathbf{u})$ . Moreover, if  $w$  is a palindrome and  $a = b$ , then  $awa$  is its *palindromic extension*. The set of all palindromic extensions of a palindrome  $w \in \mathcal{L}(\mathbf{u})$  is denoted by  $\text{Pext}(w)$ : we have

$$\text{Pext}(w) = \{awa : awa \in \mathcal{L}(\mathbf{u}), a \in \mathcal{A}\}.$$

**Theorem 4** ([4]). *Let  $\mathbf{u}$  be an infinite word that is closed under reversal. The word  $\mathbf{u}$  is rich if and only if any bispecial factor  $w$  of  $\mathbf{u}$  satisfies*

$$b(w) = \begin{cases} \#Pext(w) - 1 & \text{if } w \text{ is a palindrome;} \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

The next theorem is another characterization of rich words which will be useful.

**Theorem 5** ([13]). *A word  $\mathbf{u}$  is rich if and only if for every  $w \in \mathcal{L}(\mathbf{u})$ , any factor of  $\mathbf{u}$  containing exactly two occurrences of  $w$  or  $R(w)$ , one as a prefix and one as a suffix, is a palindrome.*

## 2.2 Words with infinitely many palindromes and morphisms

Hof, Knill and Simon studied in [22] words having infinitely many palindromes. They considered words that are constructed using morphisms of the following class:

**Definition 6.** A primitive morphism  $\varphi : \mathcal{A}^* \rightarrow \mathcal{A}^*$  belongs to the class  $P$  if there exists a palindrome  $r \in \mathcal{A}^*$  such that for any letter  $a \in \mathcal{A}$  we have that  $r$  is a prefix of  $\varphi(a)$  and  $r^{-1}\varphi(a)$  is a palindrome.

The class  $P$  is related to words with infinitely many palindromes in the following sense. Let  $\varphi$  be a morphism of class  $P$ . If  $v$  is a palindrome, then  $\varphi(v)r$  is also a palindrome. Thus, a fixed point of a morphism of class  $P$  has infinitely many palindromes.

The question is the converse: if one has a uniformly recurrent fixed point with infinitely many palindromes, can it be generated by a morphism of class  $P$ ? A discussion on the meaning of this formulation can be found in [26]. In [34], the author gave the following affirmative answer to this question in case of a binary alphabet: if a binary fixed point of a primitive morphism  $\varphi$  contains infinitely many palindromes, then  $\varphi$  or  $\varphi^2$  is conjugate to a morphism from class  $P$ . In [27], it is demonstrated that this claim cannot be generalized to a larger alphabet even under the assumption of injectivity of the morphism. In [26], the authors show that if an infinite word is a fixed point of a marked primitive morphism and contains infinitely many palindromes, then some power of the morphism has a conjugate in class  $P$ .

We now mention two classes of morphisms that are closely related to the class  $P$ . The first class is defined in [20]. A morphism  $\varphi$  is a *standard morphism of class  $P$*  (or a *standard  $P$ -morphism*) if there exists a palindrome  $r$  (possibly empty) such that for all  $x \in \mathcal{A}$ ,  $\varphi(x) = rq_x$  and  $q_x$  is a palindrome. If  $r$  is non-empty, then some (or all) of the palindromes  $q_x$  may be empty or may even take the form  $q_x = \pi_x^{-1}$  with  $\pi_x$  a proper palindromic suffix of  $r$ . A standard morphism of class  $P$  is *special* if

1. whenever  $\varphi(x)r = rq_xr$ , with  $x \in \mathcal{A}$ , occurs in some  $\varphi(y_1y_2 \cdots y_n)r$ , then this occurrence is  $\varphi(y_m)r$  for some  $m$  with  $1 \leq m \leq n$ ; and
2. all  $\varphi(x) = rq_x$  end with different letters.

Let  $\varphi$  be a special standard morphism of class  $P$ . It is related to richness in the following way: if  $\mathbf{u}$  is rich, then  $\varphi(\mathbf{u})$  is almost rich, see [20].

The second class of morphisms related to the class  $P$  is the class  $P_{ret}$ . In [5], this class was used to show relations between rich and almost rich words in general: every almost rich word is an image of a rich word by a morphism of this class. The class was also used in [10] to show that every episturmian word is an image of an Arnoux–Rauzy word. In [21], it is used to show an answer to an interpretation of the previously mentioned question of Hof, Knill and Simon.

**Definition 7.** Let  $\varphi : \mathcal{B}^* \rightarrow \mathcal{A}^*$  and  $r \in \mathcal{A}^*$  be a palindrome. We say that  $\varphi$  is of class  $P_{ret}$  (with respect to  $r$ ) if the following is true:

- $\varphi(b)r$  is a palindrome for any  $b \in \mathcal{B}$ ,
- $\varphi(b)r$  contains exactly 2 distinct occurrences of  $r$ , one as a prefix and one as a suffix, for any  $b \in \mathcal{B}$ ,
- $\varphi(b) \neq \varphi(c)$  for all  $b, c \in \mathcal{B}$ ,  $b \neq c$ .

A direct consequence of the above definition is that any morphism  $\varphi \in P_{ret}$  is injective and  $\varphi(s)r$  is a palindrome if and only if  $s \in \mathcal{B}^*$  is a palindrome. An important property of the class is also that it is closed under taking composition of morphisms, see [5].

The following example taken from [5] illustrates that there exists a morphism of class  $P_{ret}$  which maps a word having finite defect to a word with infinite defect. One of the main results of this article, Theorem 1, states that this cannot happen if such a morphism acts on an episturmian word.

**Example 8.** Let  $v_0 = \varepsilon$  and for all  $i > 0$  set

$$v_i = v_{i-1}0v_{i-1}1v_{i-1}1v_{i-1}0v_{i-1}2v_{i-1}2v_{i-1}0v_{i-1}1v_{i-1}1v_{i-1}0v_{i-1}.$$

Let  $\mathbf{v} \in \{0, 1, 2\}^{\mathbb{N}}$  be determined by the limit

$$\mathbf{v} = \lim_{i \rightarrow +\infty} v_i$$

and  $\varphi : \{0, 1, 2\}^* \rightarrow \{0, 1\}^*$  be given as follows:

$$\varphi : \begin{cases} 0 \mapsto 0100 \\ 1 \mapsto 01011 \\ 1 \mapsto 010111 \end{cases}.$$

Proposition 5.7 in [5] states that  $D(\mathbf{v}) = 0$  and  $D(\varphi(\mathbf{v})) = +\infty$ . In other words, the morphism  $\varphi$ , which is of class  $P_{ret}$ , maps a rich word to a word which is not almost rich.

The key property of  $\mathbf{v}$  in this case is that the palindromes  $v_i$  have two palindromic extensions  $1v_i1$  and  $2v_i2$  which produce (by application of  $\varphi$ ) the same palindrome  $1\varphi(v_i)0101$  but they also produce a non-palindromic complete return word to it (it is contained in  $\varphi(1v_i1v_i0v_i2v_i2)$ ). Having infinitely many non-palindromic complete return words to palindromes implies that the defect is infinite (see [20]).

### 3 Morphic images of episturmian words versus richness

In this section, we first recall and deduce some properties of episturmian words that will be needed later. A proof of Theorem 1 is given in the second subsection. The last subsection contains a proof of Theorem 3.

#### 3.1 Properties of episturmian words

For basic reference on episturmian words, the reader can refer to [17, 24] or survey papers [19, 7].

As already mentioned, an infinite word  $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$  is *episturmian* if for any  $n$  there exists at most one left special factor of length  $n$  and  $\mathbf{u}$  is closed under reversal. If for any  $n$  there exists exactly one left special factor of length  $n$  with  $\#\mathcal{A}$  left extensions, then  $\mathbf{u}$  is a  $k$ -ary *Arnoux–Rauzy* word with  $k = \#\mathcal{A}$ .

The most important examples of episturmian words can be constructed using the palindromic closure. The standard word with the directive sequence  $(\delta_i)_{i=0}^{+\infty} \in \mathcal{A}^{\mathbb{N}}$  and with the seed  $w_0 = \varepsilon$ , i.e., the word  $\mathbf{u}$  such that

$$\mathbf{u} = \lim_{n \rightarrow +\infty} w_n \quad \text{where } w_n = (w_{n-1}\delta_{n-1})^R,$$

is an episturmian word over  $\mathcal{A}$  and it is called *standard episturmian*. The importance follows from the fact that given any episturmian word  $\mathbf{u}$ , there exists a unique standard episturmian word with the same language. Since richness is a property of language, when studying it, we can restrict ourselves to the standard episturmian words which are characterized by their directive sequence. A standard episturmian word can be recognized by looking at its prefixes: an episturmian word is standard if and only if each its prefix is a left special factor.

A basic property of any episturmian word  $\mathbf{u}$  is that one letter is *separating*: it occurs in every factor of length 2. The separating letter of  $\mathbf{u}$  is the first letter of the directive sequence  $\Delta = \delta_1\delta_2\delta_3\dots$  of the corresponding standard episturmian word. Denote  $\ell$  the least integer such that  $\delta_1^\ell$  is not a prefix of  $\Delta$ . Then the word  $x\delta_1^k y$  with  $x, y \neq \delta_1$  is a factor of the episturmian word  $\mathbf{u}$  only if  $k = \ell - 1$  or  $k = \ell$ . Moreover, the word  $\delta_1^\ell$  is a factor of the episturmian word  $\mathbf{u}$  if and only if the letter  $\delta_1$  occurs at least once in the sequence  $\delta_{\ell+1}\delta_{\ell+2}\delta_{\ell+3}\dots$ .

Any palindromic prefix of a standard episturmian word is equal to  $w_n$  for some  $n \in \mathbb{N}$ . The complete return words of  $w_n$  are described by Theorem 4.4 of [25]. Denote by  $\mathcal{F}$  the set  $\{\delta_m : m \geq n\}$ . Every complete return word of  $w_n$  equals to the palindromic closure

$$(w_n x)^R = q_x w_n$$

where  $x \in \mathcal{F}$ . The word  $q_x$  is the corresponding return word. Thus, the palindromic prefix  $w_n$  has exactly  $\#\mathcal{F}$  return words.

Bispecial factors play a crucial role in the study of the language of an infinite word. We give essential properties of bispecial factors of an episturmian word. Any bispecial



factor  $w$  of an episturmian word  $\mathbf{u}$  is a palindrome; moreover there exists  $n \in \mathbb{N}$  such that  $w = w_n$ , where  $w_n$  is a palindromic prefix of the corresponding standard episturmian word. Denoting again  $\mathcal{F} = \{\delta_m : m \geq n\}$ , the set of all non-palindromic both-sided extensions of the bispecial factor  $w$  satisfies

$$\{xwy \in \mathcal{L}(\mathbf{u}) : x, y \in \mathcal{A}, x \neq y\} = \{awx \in \mathcal{L}(\mathbf{u}) : x \in \mathcal{F} \setminus \{a\}\} \cup \{xwa \in \mathcal{L}(\mathbf{u}) : x \in \mathcal{F} \setminus \{a\}\}, \quad (3)$$

where  $a \in \mathcal{F}$  is the unique letter such that  $awx \in \mathcal{L}(\mathbf{u})$  and  $xwa \in \mathcal{L}(\mathbf{u})$  for some letter  $x \in \mathcal{A}$ . If  $\mathbf{u}$  is Arnoux–Rauzy, then  $awa \in \mathcal{L}(\mathbf{u})$  (see [16]) (and  $\mathcal{F} = \mathcal{A}$ ); if  $\mathbf{u}$  is not Arnoux–Rauzy, then it depends on  $w$  whether  $awa \in \mathcal{L}(\mathbf{u})$  (and there exist a bispecial factor  $w$  such that  $awa \notin \mathcal{L}(\mathbf{u})$ ).

The next lemmas give an insight on the structure of return words of an episturmian word.

**Lemma 9.** *Let  $\mathbf{u}$  be an episturmian word over  $\mathcal{A}$  and  $w \in \mathcal{L}(\mathbf{u})$  be a factor. Let  $r_1, r_2, \dots, r_s$  be the list of all distinct return words of  $w$  in  $\mathbf{u}$ . Define a morphism  $\Psi$  over  $\mathcal{E} = \{1, 2, \dots, s\}$  by the rule  $k \mapsto r_k$  for all  $k \in \mathcal{E}$ . There exists an episturmian word  $\mathbf{v}$  over  $\mathcal{E}$  such that  $\mathbf{u} = g\Psi(\mathbf{v})$  for some finite word  $g$ .*

*Proof.* First, let us assume that  $w$  is a bispecial factor of  $\mathbf{u}$ . If  $\mathbf{u}$  is an Arnoux–Rauzy word, then  $\#\mathcal{E} = \#\mathcal{A}$  and the claim follows directly from the proof of Theorem 1 in [10] and from Theorem 3 ibidem. In case  $\#\mathcal{E} < \#\mathcal{A}$ , the proof is analogous and is left to the reader.

Suppose  $w$  is not bispecial and it can be extended in a unique way to the shortest bispecial factor  $b = uvv$  in which it occurs. A factor  $r$  is a return word of  $b$  if and only if  $u^{-1}ru$  is a return word of  $w$ . Thus the morphism  $\Psi$  (defined using the return words of  $w$ ) is a conjugate morphism of the morphism defined using the return words of  $b$  and we can use the validity of the statement for the bispecial factor  $b$ .

The last case,  $w$  is not bispecial and it cannot be extended to a bispecial factor, is trivial as  $w$  has only one return word and thus  $\#\mathcal{E} = 1$ .  $\square$

The word  $\mathbf{v}$  from Lemma 9 captures the structure of return words of the factor  $w$  in  $\mathbf{u}$ . Such a word was also studied in [18] and we keep the same terminology: we say that the word  $\mathbf{v}$  from the previous lemma is a *derivated word of  $\mathbf{u}$  with respect to the factor  $w$*  and its corresponding morphism is  $\Psi$ . By comparing the definitions we obtain that if  $w$  is a palindrome, then  $\Psi \in P_{ret}$ .

**Lemma 10.** *Let  $\mathbf{u}$  be a standard episturmian word  $\mathbf{u}$  over the alphabet  $\mathcal{A} = \{0, 1, 2\}$  with the directive sequence  $\delta_1\delta_2\delta_3\dots$ , where  $\delta_1 = 0$ . Denote by  $\ell$  the least integer such that  $0^\ell$  is not a prefix of the directive sequence.*

1. *If  $w = 1$  or  $w = 2$ , then no two return words of  $w$  have the same length.*
2. *If there exists  $j$  such that  $j > \ell$  and  $\delta_j = 0$ , then  $w = 0^\ell$  is a factor of  $\mathbf{u}$  and no two return words of  $w$  have the same length.*



*Proof.* If  $w$  has exactly 1 return word, then the claim is trivially satisfied.

Suppose  $w$  has more than 1 return word. Let  $b = uvv$  be the shortest bispecial factor containing  $w$ . Any return word of  $w$  has the form  $u^{-1}ru$ , where  $r$  is a return word of  $b$ . Therefore, it is enough to prove the statement only for bispecial factors. Since  $\mathbf{u}$  has a prefix  $0^{\ell-1}$  but not  $0^\ell$ , the bispecial factor  $b$  contains at least two distinct letters.

As already mentioned, Theorem 4.4 of [25] describes the return words of bispecial factors. In particular, any complete return word of  $b$  equals to the palindromic closure of  $bx$  for some letter  $x \in \mathcal{A}$ . Moreover, since  $\mathbf{u}$  is standard, each bispecial factor  $b$  of  $\mathbf{u}$  is a palindromic prefix of  $\mathbf{u}$ . Since the prefixes of the word are constructed using the palindromic closure and the directive sequence, the palindromic closures  $(bx)^R$  and  $(by)^R$  for distinct letters  $x$  and  $y$  have the same length if and only if neither  $x$  nor  $y$  occurs in  $b$  (see [23]). As  $b$  contains at least two distinct letters, this cannot happen on the ternary alphabet.  $\square$

*Remark 11.* As one can see from the proof of the last lemma, for alphabets having cardinality greater than 3 the claim of the last lemma does not hold. Let us suppose that  $\mathbf{u}$  is a standard episturmian word with its directive sequence  $\Delta = 01023\dots$ . We have

$$\mathbf{u} = 010010201001030100102010010\dots$$

Clearly,  $\ell = 2$ . The factors 0010201 and 0010301 are return words of  $w = 0^\ell = 00$  and they are of the same length.

**Lemma 12.** *Let  $\mathbf{u}$  be an episturmian word over  $\mathcal{A}$ . Let  $\mathcal{A}' \subset \mathcal{A}$  and  $xpy$  be a factor of  $\mathbf{u}$  such that  $x, y \in \mathcal{A}'$  and  $p$  does not contain any letter from  $\mathcal{A}'$ . The word  $p$  is a palindrome.*

*Proof.* We can suppose without loss of generality that the word  $\mathbf{u}$  is standard episturmian. We order the elements of  $\mathcal{A}'$  according to their first occurrence in the directive sequence, i.e.,

$$\mathcal{A}' = \{x_1, x_2, \dots, x_n\},$$

where  $x_i$  appears for the first time in the directive sequence of  $\mathbf{u}$  before the first occurrence of  $x_j$  if and only if  $i < j$ .

Suppose  $x = x_1$ . Let  $p_1$  be the longest prefix of  $\mathbf{u}$  that does not contain  $x_1$  (or equivalently, the longest prefix of  $\mathbf{u}$  over  $\mathcal{A} \setminus \mathcal{A}'$ ). Since  $\mathbf{u}$  is a standard word constructed using the palindromic closure operators, the word  $p_1$  is a palindrome (it may be empty). The shortest palindromic prefix containing  $x_1$  is  $p_1x_1p_1$ . Using Theorem 4.4 of [25] we have that any complete return word of  $p_1x_1p_1$  has the form  $(p_1x_1p_1z)^R$  for some  $z \in \mathcal{A}$ . If  $p_1x_1p_1$  is bispecial, it is the shortest bispecial containing  $x_1$ ; if it is not bispecial, then  $\mathbf{u}$  is periodic. In both cases, any word  $x_1py$  is a prefix of a complete return word of  $p_1x_1p_1$  without the leading  $p_1$ . If  $z \in \mathcal{A}'$ , then  $y = z$  and  $xpy = x_1p_1z$ , i.e.,  $p = p_1$  is a palindrome. If  $z \notin \mathcal{A}'$ , then  $y = x_1$  and  $xpy = p_1^{-1}(p_1x_1p_1z)^Rp_1^{-1}$ , i.e.,  $p$  is again a palindrome.

Take  $j > 1$  and suppose  $x = x_j$ . Let  $p_j$  be the longest prefix of  $\mathbf{u}$  that does not contain  $x_j$ . The shortest palindromic prefix of  $\mathbf{u}$  containing  $x_j$  is the word  $p_jx_jp_j$ . If  $p_jx_jp_j$  is bispecial, it is the shortest bispecial factor containing  $x_j$ ; if it is not bispecial, then  $\mathbf{u}$

is periodic. Both cases imply that the any complete return word of  $x_j$  starts with  $x_j p_j$ . Since  $x_j p_j = x_j p_1 x_1 s$  for some word  $s$ , there is just one word of the desired form, precisely  $x_j p_1 x_1$ , i.e.,  $y = x_1$ , and the proof is finished.  $\square$

When dealing with images by a morphism a useful notion is an ancestor. Suppose that  $w$  is a factor of  $\Psi(\mathbf{v})$  for some infinite word  $\mathbf{v}$  and morphism  $\Psi$ . A factor  $e_0 e_1 \cdots e_n \in \mathcal{L}(\mathbf{v})$  is an *ancestor* by  $\Psi$  of  $w$  if  $\Psi(e_0 e_1 \cdots e_n)$  contains  $w$  but neither  $\Psi(e_1 \cdots e_n)$  nor  $\Psi(e_0 e_1 \cdots e_{n-1})$  contains  $w$ .

The following lemma can be considered as a generalization of Theorem 5. The lemma uses an additional notion. Let us fix a factor  $w$  of an episturmian word  $\mathbf{u}$  over the alphabet  $\mathcal{A}$  and a subset  $\mathcal{E} \subset \mathcal{A}$ . A both-sided extension  $xwy$  of  $w$  in  $\mathbf{u}$  is called  $\mathcal{E}$ -*extension* if  $x$  and  $y$  belong to  $\mathcal{E}$ .

**Lemma 13.** *Let  $w$  be a palindromic factor of an episturmian word  $\mathbf{u} \in \mathcal{A}^{\mathbb{N}}$  and  $a \in \mathcal{A}$  be the unique letter satisfying  $awx, xwa \in \mathcal{L}(\mathbf{u})$  for some letter  $x$ . Let  $\mathcal{E} \subset \mathcal{A}$  such that  $a \in \mathcal{E}$ . If  $xuy$  is a factor of  $\mathbf{u}$  containing exactly two  $\mathcal{E}$ -extensions of  $w$  — one as its prefix and one as its suffix — then  $u$  is a palindrome.*

*Proof.* If  $w$  is not a bispecial factor, then the only both-sided extension of  $w$  is  $awa$ . It means that  $xuy = aua$  is a complete return word of the palindrome  $awa$  and thus, using Theorem 5, the word  $u$  is a palindrome.

For a bispecial factor  $w$ , we denote by  $n$  the number of occurrences of the factor  $w$  in  $u$ .

If  $n = 2$ , then  $u$  is a complete return word of  $w$  and, since  $\mathbf{u}$  is rich, using Theorem 5  $u$  is a palindrome.

If  $n > 2$ , we first show that the factor  $xuy$  begins with  $xwa$  and ends with  $awy$ . Suppose that  $xuy$  begins with  $xwy$ ,  $y \neq a$ . There is only one complete return word of  $w$  beginning with  $wy$ , it ends with  $yw$  and it is always followed by the letter  $a$ , thus we have a contradiction to  $n > 2$ .

We deal with two subcases:

1)  $x = y$ .

Since the word  $xuy$  has a prefix  $p = xwa$ , it has a suffix  $R(p) = awx$ . The factor  $xuy = xux$  does not contain further occurrences of  $p$  and  $R(p)$ . According to Theorem 5, such a factor is a palindrome in any rich word, in particular in the episturmian word  $\mathbf{u}$ , as we want to show.

2)  $x \neq y$ .

Let  $\mathcal{B}$  be the set of letters such that for all  $z \in \mathcal{B}$  the word  $wz$  is a right extension of  $w$ . For all  $z \in \mathcal{B}$ , denote  $r_z$  the return word of  $w$  which has a prefix  $wz$ . Define a morphism  $\Psi : \mathcal{B}^* \rightarrow \mathcal{A}^*$  by setting  $\Psi(z) = r_z$  for all  $z \in \mathcal{B}$ . Using Lemma 9, let  $\mathbf{v} \in \mathcal{B}^{\mathbb{N}}$  be the episturmian word which is a derivated word of  $\mathbf{u}$  with respect to the factor  $w$  such that  $\Psi$  is the corresponding morphism. The ancestor by  $\Psi$  of the factor  $xuy$  is a factor  $xasay \in \mathcal{L}(\mathbf{v})$  with  $s$  containing only letters from  $\{z : z = a \text{ or } z \notin \mathcal{E}\}$  and moreover, in any pair of neighboring letters in  $asa$  exactly one of the letters is  $a$ , i.e.,  $aa$  does not occur in  $asa$ .

If  $x, y \neq a$ , then we apply Lemma 12 with  $\mathcal{A}' = \mathcal{E} \setminus \{a\}$ . We get that  $asa$  is a palindrome and thus  $u = \Psi(asa)w$  is a palindrome as well.

If  $x = a$  and  $y \neq a$ , then since  $a$  is the separating letter in  $\mathbf{v}$ , we have that  $xasay = aasay$  is an image of the morphism  $\sigma_a$  defined by  $a \mapsto a$  and  $b \mapsto ab$  for any  $b \neq a$ . Thus  $aasay = \sigma_a(a\tilde{s}y)$  where  $a\tilde{s}y$  is a factor of an episturmian word and  $\tilde{s}$  is produced from  $s$  by erasing all letters  $a$  in it. Moreover, the word  $\tilde{s}$  does not contain letters from  $\mathcal{E}$ . Applying Lemma 12 with  $\mathcal{A}' = \mathcal{E}$  we get that  $\tilde{s}$  is a palindrome. Since  $\sigma_a$  and  $\Psi$  belong to the class  $P_{ret}$ , their composition  $\Psi\sigma_a$  is in  $P_{ret}$  with respect to  $\Psi(a)w$ . Consequently,  $u = \Psi(\sigma_a(\tilde{s})a)w$  is a palindrome.  $\square$

*Remark 14.* The morphism  $\sigma_a$  is one of *episturmian morphisms*, see [24]. For instance, episturmian morphisms can be used to construct episturmian words.

### 3.2 Proof of Theorem 1

Besides the proof of Theorem 1 which concerns images of episturmian words under a morphism of class  $P_{ret}$ , we also provide in this section a description of the bilateral order of long bispecial factors in these infinite words. This result is needed also to prove Theorem 3 in the next section.

**Lemma 15.** *Let  $\mathbf{u}$  be an episturmian word and let  $\pi : \mathcal{A}^* \rightarrow \mathcal{B}^*$  a morphism of class  $P_{ret}$  with respect to the palindrome  $r$ . If  $w \in \mathcal{L}(\pi(\mathbf{u}))$  is a bispecial factor containing at least one occurrence of  $r$ , then its bilateral order  $b(w)$  is  $-1$  or  $0$ , and satisfies (2).*

*Proof.* Let  $w \in \mathcal{L}(\pi(\mathbf{u}))$  be a bispecial factor containing  $r$ . Let us denote  $sr$  and  $rp$  the prefix and the suffix of  $w$  containing the factor  $r$  exactly once. As  $\pi$  belongs to the class  $P_{ret}$ , there exists a unique factor  $v \in \mathcal{L}(\mathbf{u})$  such that  $w = s\pi(v)rp$ .

Let  $\{z_1, \dots, z_t\}$  be all the letters such that  $wz_i$  is a right extension of  $w$ . We denote by  $R_i = \{b \in \mathcal{A} : \pi(v)rpz_i \text{ is a prefix of } \pi(vb)r \text{ and } vb \in \mathcal{L}(\mathbf{u})\}$ . This set is well defined due to the choice of  $p$  and  $v$ . Since  $\pi$  is of class  $P_{ret}$ , it follows that  $R_i \cap R_j = \emptyset$  if  $i \neq j$ . Thus,  $v$  is right special.

Let  $\{y_1, \dots, y_\ell\}$  be all the letters such that  $y_iw$  is a left extension of  $w$ . Analogously to  $R_i$ , we define the sets  $L_i$ . We conclude that  $v$  is left special, thus it is bispecial.

Denote by  $a$  the unique letter such that  $avx \in \mathcal{L}(\mathbf{u})$  and  $xva \in \mathcal{L}(\mathbf{u})$  for some  $x \in \mathcal{A}$ . Next, we show that there exists an index  $i$  such that  $a \in R_i$ . Suppose the contrary: for all  $i$  and  $b \in R_i$  we have that  $wz_i$  is a factor of  $\pi(cvb)$  for some letter  $c$  (possibly depending on  $b$ ). Using (3), since  $b \neq a$ , it implies that  $c = a$ , i.e.,  $v$  is not special — a contradiction. Let without loss of generality  $a \in R_1$ . Analogously, we have  $a \in L_1$ .

Using the property (3) for  $v$ , we have that if  $y_jwz_k \in \mathcal{L}(\pi(\mathbf{u}))$ , then  $k > 1$  implies  $j = 1$  and symmetrically  $j > 1$  implies  $k = 1$ . Thus we have

$$\{ywz \in \mathcal{L}(\pi(\mathbf{u}))\} \setminus \{y_1wz_1\} = \{y_1wz_k : k > 1\} \cup \{y_kwz_1 : k > 1\}.$$

To calculate the bilateral order of  $w$ , it remains to see if  $y_1wz_1$  is its both-sided extension or not: if  $y_1wz_1 \in \mathcal{L}(\pi(\mathbf{u}))$ , then  $b(w) = 0$ ; otherwise  $b(w) = -1$ .

We first show that if  $R_1 = L_1 = \{a\}$ , then  $w$  is a palindrome. Since  $R_1 = \{a\}$ , there is no letter  $b \in R_i$  for some  $i$  such that the longest common prefix of  $\pi(b)r$  and  $\pi(a)r$  is longer than  $|rp|$ . Similarly, there is no letter  $c \in L_j$  for some  $j$  such that the longest common suffix of  $\pi(c)$  and  $\pi(a)$  is longer than  $|s|$ . This implies  $\bigcup R_i = \bigcup L_i$ , and  $s = R(rp)$ , i.e.,  $w$  is a palindrome. Moreover, we obtain  $t = \ell$ , and for all  $i$  there exists  $j$  such that  $R_i = L_j$ . It implies that  $y_i w z_j$  is a palindrome if and only if  $i = j = 1$ .

We have that  $y_1 w z_1 \in \mathcal{L}(\pi(\mathbf{u}))$  if and only if  $ava \in \mathcal{L}(\mathbf{u})$  or  $R_1 \neq \{a\}$  or  $L_1 \neq \{a\}$ .

Thus, if  $y_1 w z_1 \notin \mathcal{L}(\pi(\mathbf{u}))$ , i.e.,  $b(w) = -1$ , then  $R_1 = L_1 = \{a\}$ , which implies that  $w$  is a palindrome and  $w$  has no palindromic extension. Therefore,  $b(w) = \# \text{Pext}(w) - 1 = -1$  and (2) is satisfied.

The last case is  $y_1 w z_1 \in \mathcal{L}(\pi(\mathbf{u}))$ . If  $w$  is a palindrome, since  $y_1 w z_1$  is the only palindromic extension, (2) is again satisfied. If  $w$  is not a palindrome, we have  $b(w) = 0$ , and (2) is also satisfied.  $\square$

*Proof of Theorem 1.* Let  $\pi$  be a morphism from  $P_{ret}$  with respect to  $r$ . Since  $\mathbf{u}$  is a derivated word of  $\pi(\mathbf{u})$  with respect to  $r$ , then  $\mathbf{u}$  is periodic if and only if  $\pi(\mathbf{u})$  is periodic.

Suppose  $\mathbf{u}$  is periodic. Thus,  $\pi(\mathbf{u})$  is also periodic. In Theorem 4 of [9], it is shown that an infinite periodic word contains infinitely many palindromes if and only if it is equal to  $(pq)^\omega$  for some palindromes  $p$  and  $q$ . Since  $\mathbf{u}$  contains infinitely many palindromes,  $\pi(\mathbf{u})$  contains infinitely many palindromes and we can write  $\pi(\mathbf{u}) = (pq)^\omega$  for some palindromes  $p$  and  $q$  from  $\mathcal{L}(\pi(\mathbf{u}))$ . It can be easily seen that such a word has finite defect.

Suppose  $\mathbf{u}$  is aperiodic. As shown in [5], a uniformly recurrent infinite word  $\pi(\mathbf{u})$  containing infinitely many palindromes has finite defect if and only if there exists an integer  $M$  such that any complete return word of any palindrome  $w$  of length at least  $M$  is palindromic. As for each palindrome  $p \in \mathcal{L}(\mathbf{u})$ , the word  $\pi(p)r$  is a palindrome, and moreover a factor of  $\pi(\mathbf{u})$ , it remains to verify palindromicity of complete return words to palindromic factors of length  $N$  such that  $N \geq M = 2 \max\{|\pi(a)r| : a \in \mathcal{A}\}$ . Let  $w$  be a palindromic factor of  $\pi(\mathbf{u})$  such that  $|w| \geq M$ .

First, suppose that  $w$  is a bispecial factor. The word  $w$  contains  $r$  and according to the proof of Lemma 15 there exists a bispecial palindromic factor  $v$ , a letter  $a$  and a factor  $s$  such that for some letter  $x$  we have

$$w = s\pi(v)rR(s), \quad avx \in \mathcal{L}(\mathbf{u}), \quad xva \in \mathcal{L}(\mathbf{u}) \quad \text{and} \quad sr \text{ is a suffix of } \pi(a)r.$$

Let  $f$  be a complete return word of  $w$ . Clearly,  $f = s\pi(u)rR(s)$  for some factor  $u \in \mathbf{u}$  and  $v$  is a prefix and a suffix of  $u$ . If  $s$  is the empty word  $\varepsilon$ , then  $u$  is a complete return word of  $v$  in  $\mathbf{u}$ . Since  $u$  is a palindrome, the factor  $f = \pi(u)r$  is a palindrome as well.

If  $s \neq \varepsilon$ , then we apply Lemma 13 with  $\mathcal{E} = \{b \in \mathcal{A} : sr \text{ is a suffix of } \pi(b)r\}$ . According to our notation any ancestor of  $w$  has the form  $xvy$ , where  $x, y \in \mathcal{E}$ . Moreover, any ancestor of the complete return word  $f = s\pi(u)rR(s)$  equals  $xuy$  with  $x, y \in \mathcal{E}$  and the factor  $xuy$  has only two occurrences of  $\mathcal{E}$ -extensions of the factor  $v$ . According to Lemma 13, the factor  $u$  is a palindrome and thus  $\pi(u)r$  is a palindrome as well. It implies that the complete return word  $f = s\pi(u)rR(s)$  is a palindrome, too.

To complete the proof, we need to discuss the case when  $w$  is not bispecial. Since  $\pi(\mathbf{u})$  is closed under reversal, such a palindrome  $w$  has a unique both-sided extension and this extension is palindromic. It implies that the shortest bispecial factor in which the palindrome  $w$  occurs is a bispecial palindrome, say  $qwR(q)$ . Moreover,  $u$  is a complete return word of  $w$  if and only if  $qwR(q)$  is a complete return word of  $u$ . Since  $qwR(q)$  is palindrome,  $u$  is palindrome as well.  $\square$

*Remark 16.* Given an episturmian word  $\mathbf{u}$  and a morphism  $\pi$  of class  $P_{ret}$ , the defect of  $\pi(\mathbf{u})$  can be nonzero. It suffices to choose  $\pi$  such that for some letter  $a \in \mathcal{A}$  the defect of  $\pi(a)$  is nonzero, which is possible in general.

For instance, let  $\mathbf{u}$  be the Fibonacci word, i.e., the fixed point of the morphism given by  $0 \mapsto 01$  and  $1 \mapsto 0$ . The Fibonacci word is Sturmian, i.e., an episturmian word. Let  $\pi \in \mathcal{P}_{ret}$  be determined by

$$0 \mapsto 110100110010 \quad \text{and} \quad 1 \mapsto 1.$$

(The morphism  $\pi$  is of class  $P_{ret}$  with respect to the palindrome  $r = 11$ .) We have  $D(\pi(0)) = 1$ , thus  $D(\pi(\mathbf{u})) \geq 1$ . In fact,  $D(\pi(\mathbf{u})) = 2$ .

### 3.3 Proof of Theorem 3

Before proving Theorem 3, we need two lemmas concerned with images of episturmian words by morphisms of a special form.

**Lemma 17.** *Let  $\mathbf{u}$  be an episturmian word over  $\mathcal{A}$  and  $\pi : \mathcal{A} \rightarrow \{A, B\}$  a morphism of class  $P_{ret}$  such that  $\pi(x) = AB^{q_x}$  for all  $x$ . The word  $\pi(\mathbf{u})$  is rich.*

*Proof.* Any morphism from class  $P_{ret}$  is injective, thus  $q_x \neq q_y$  for any  $x, y \in \mathcal{A}, x \neq y$ . We will show that if  $w$  is a bispecial factor of  $\pi(\mathbf{u})$ , then its bilateral order  $b(w)$  satisfies (2).

If the bispecial factor  $w$  contains the letter  $A$ , the claim follows from Lemma 15. Suppose  $w$  does not contain any occurrence of the letter  $A$ , i.e.,  $w = B^k$  with  $k < \max_{x \in \mathcal{A}} q_x$ .

The set of both-sided extensions of  $w$  clearly contains  $AB^{k+1}$  and  $B^{k+1}A$ . Thus the set of both-sided extensions of  $w$  has cardinality  $2 + \#\text{Pext}(w)$ . Since  $\#\text{Rext}(w) = 2 = \#\text{Lext}(w)$ , according to equation (1) we have  $b(w) = \#\text{Pext}(w) - 1$ .

Since (2) is satisfied for any bispecial factor of  $\pi(\mathbf{u})$ , the richness of  $\pi(\mathbf{u})$  follows from Theorem 4.  $\square$

**Lemma 18.** *Let  $\mathbf{u}$  be a ternary episturmian word over  $\{0, 1, 2\}$  with the separating letter 0 and let  $\pi : \mathcal{A} \rightarrow \{A, B\}$  be a morphism such that  $\pi(0) = A$  and  $\pi(x) = AB$  for all  $x \neq 0$ . The word  $\pi(\mathbf{u})$  is rich.*

*Proof.* Let  $\ell$  be the minimal integer such  $0^\ell$  is not a prefix of the directive sequence of the standard episturmian word having the same language as  $\mathbf{u}$ . It means that any two non-zero letters of  $\mathbf{u}$  are separated by the block  $0^\ell$  or  $0^{\ell-1}$ .

If  $0^\ell$  does not belong to  $\mathcal{L}(\mathbf{u})$ , then  $\pi(\mathbf{u}) = (A^\ell B)^\omega$ , which is a rich word.

Let us suppose that  $0^\ell$  occurs in  $\mathbf{u}$ . Using Lemma 9, let  $\mathbf{v} \in \mathcal{B}^\mathbb{N}$  be an episturmian word that is a derivated word of  $\mathbf{u}$  with respect to the factor  $0^\ell$  and let  $\Psi$  be the corresponding morphism. Thus, for any letter  $b \in \mathcal{B}$  the image  $\Psi(b)$  is a return word of  $0^\ell$  in  $\mathbf{u}$ . Using Lemma 10 one can see that the morphism  $\pi\Psi$  is of class  $P_{ret}$  with respect to the palindrome  $r = A^{\ell+1}$ .

We will show that if  $w$  is a bispecial factor of  $\pi(\mathbf{u})$ , then the bilateral order  $b(w)$  satisfies (2). If the bispecial factor  $w$  contains  $r$ , the claim follows from Lemma 15.

Suppose  $w$  does not contain any occurrence of the word  $r$ . Since any occurrence of the letter  $B$  in  $\pi(\mathbf{u})$  is followed by block  $A^\ell$  or  $A^{\ell+1}$  one can see that the longest factor (not necessarily bispecial) of  $\pi(\Psi(\mathbf{v}))$  which does not contain  $r = A^{\ell+1}$  is of the form  $A^\ell(BA^\ell)^k$  for some integer  $k$ . Thus the bispecial factor  $w$  is of the form  $w = A^\ell(BA^\ell)^s$ , with  $s < k$ . Since  $A^\ell(BA^\ell)^k$  is a factor, we can see that  $AwB$  and  $BwA$  are factors of  $\pi(\mathbf{u})$ . Since  $w$  is a palindrome, we have  $b(w) = \#Pext(w) - 1$  and consequently  $w$  satisfies (2) as well.

Since (2) is satisfied for any bispecial factor of  $\pi(\mathbf{u})$ , Theorem 4 implies that  $\pi(\mathbf{u})$  is rich.  $\square$

*Proof of Theorem 3.* Without loss of generality let us suppose that  $\mathcal{A} = \{0, 1, 2\}$  and 0 is the separating letter of  $\mathbf{u}$ . Let  $x \in \mathcal{A}$  be the letter such that  $\zeta(x) \neq \zeta(y)$  for all  $y \in \mathcal{A}$  such that  $y \neq x$ .

Suppose  $x = 0$ . Let  $\mathbf{v}$  be a derivated word of  $\mathbf{u}$  with respect to the factor 0 and let  $\Psi$  be the corresponding morphism. Since the return words of  $x = 0$  are 0 and  $0y$  for all  $y \neq 0$ , we may choose  $\Psi$  such that  $\Psi(0) = 0$ . It implies that the morphism  $\pi = \zeta\Psi$  and the word  $\mathbf{v}$  satisfy the assumptions of Lemma 18 which implies that  $\pi(\mathbf{v}) = \zeta(\mathbf{u})$  is rich.

Suppose  $x \neq 0$ , i.e.,  $x$  is not the separating letter in  $\mathbf{u}$ . Let again  $\mathbf{v}$  be a derivated word of  $\mathbf{u}$  with respect to the word  $x$  and let  $\Psi$  be the corresponding morphism. Using Lemma 10, one can see that  $\pi = \zeta\Psi$  is of the form as in the assumptions Lemma 17 which implies that  $\pi(\mathbf{v}) = \zeta(\mathbf{u})$  is rich.  $\square$

## 4 Application: construction of $H$ -rich words

A generalization of the notion of richness was introduced in [29]. As already mentioned, Theorems 1 and 3 and results of [30] may be used to produce examples of words that are rich in this generalized sense. In order to give these examples, and to give the definition of generalized richness on binary alphabets, we need a few more definitions.

Let  $E$  be the antimorphism over  $\{0, 1\}$  which exchanges the two letters, i.e., for instance we have  $E(011) = E(1)E(1)E(0) = 001$ . Let  $\Psi \in \{R, E\}$ . If  $p = \Psi(p)$ , the word  $p$  is a  $\Psi$ -palindrome. The  $\Psi$ -palindromic complexity of an infinite word  $\mathbf{u}$  is the mapping  $\mathcal{P}_\mathbf{u}^\Psi : \mathbb{N} \rightarrow \mathbb{N}$ , defined by  $\mathcal{P}_\mathbf{u}^\Psi(n) = \#\{p \in \mathcal{L}(\mathbf{u}) : p = \Psi(p), |p| = n\}$ .

Let  $H = \{R, E, RE, Id\}$ . For binary words closed under all elements of  $H$ , it is shown in [29] that

$$\mathcal{C}_\mathbf{u}(n+1) - \mathcal{C}_\mathbf{u}(n) + 4 \geq \mathcal{P}_\mathbf{u}^R(n+1) + \mathcal{P}_\mathbf{u}^R(n) + \mathcal{P}_\mathbf{u}^E(n+1) + \mathcal{P}_\mathbf{u}^E(n) \quad \text{for every } n \geq 1. \quad (4)$$



When equality is reached for every  $n \geq 1$ , we say that the word is *H-rich*. When the equality holds except for finitely many  $n$ , then the word is *almost H-rich*.

In [29], the notion of generalized richness is defined for arbitrary alphabet and a finite group of morphisms and antimorphisms  $G$  containing at least one antimorphism. In this context, the “classical” richness defined in the previous sections is for  $G = \{\text{Id}, R\}$ , and it coincides with (almost)  $\{\text{Id}, R\}$ -richness, or shortly (almost)  $R$ -richness.

The following construction of  $H$ -rich and almost  $H$ -rich words uses the operation  $S$  acting over words over the alphabet  $\{0, 1\}$  and defined by  $S(u_0 u_1 u_2 \dots) = v_1 v_2 v_3 \dots$ , where  $v_i = u_{i-1} + u_i \pmod 2$ . It also uses the two following results of [30].

**Proposition 19** ([30]). *Let  $\mathbf{v} = S(\mathbf{w}) \in \{0, 1\}^{\mathbb{N}}$  be uniformly recurrent. If  $\mathcal{L}(\mathbf{v})$  contains infinitely many  $R$ -palindromes centered at the letter 1 and infinitely many  $R$ -palindromes not centered at the letter 1, then  $\mathbf{w}$  is closed under all elements of  $H$ .*

**Theorem 20** ([30]). *Let  $\mathbf{w} \in \{0, 1\}^{\mathbb{N}}$  and let  $\mathbf{w}$  be closed under all elements of  $H = \{\text{Id}, E, R, ER\}$ . Then  $\mathbf{w}$  is  $H$ -rich (resp. almost  $H$ -rich) if and only if  $S(\mathbf{w})$  is  $R$ -rich (resp. almost  $R$ -rich).*

In order to produce an almost  $H$ -rich word using the last two claims, we need to find a suitable uniformly recurrent binary almost  $R$ -rich word  $\mathbf{v}$  which contains infinitely many  $R$ -palindromes centered at the letter 1 and infinitely many  $R$ -palindromes not centered at the letter 1. We will apply a well chosen morphism  $\pi \in P_{ret}$  on an episturmian word to produce such a word  $\mathbf{v}$ .

Let us recall that any Arnoux–Rauzy word over  $\mathcal{A}$  has infinitely many palindromes centered at  $a$  for each  $a \in \mathcal{A} \cup \{\varepsilon\}$ , see [16]. Consider  $\pi : \mathcal{A}^* \rightarrow \mathcal{B}^*$  a morphism from the class  $P_{ret}$  with respect to a palindrome  $r$ .

- If  $v$  is a palindrome of even length, the palindrome  $\pi(v)r$  is centered at the same letter as  $r$ .
- If  $v$  is a palindrome of odd length, centered at  $a \in \mathcal{A}$ , then  $\pi(v)r$  is centered at the same letter as  $\pi(a)r$ .

Thus, when we have an Arnoux–Rauzy word  $\mathbf{u}$  over  $\mathcal{A}$ , we can always pick a morphism  $\pi : \mathcal{A}^* \rightarrow \{0, 1\}^*$  such that  $\mathbf{v} = \pi(\mathbf{u})$  has infinitely many palindromes centered at 1 and infinitely many palindromes not centered at 1. Using Theorem 1, we have that  $\mathbf{v}$  is almost  $R$ -rich. Since  $\mathbf{u}$  is uniformly recurrent, if  $\mathbf{w}$  is a word such that  $\mathbf{v} = S(\mathbf{w})$ , then it is closed under all elements of  $H$  by Proposition 19. Theorem 20 implies that  $\mathbf{w}$  is almost  $H$ -rich.

Using Theorem 3, we can produce  $H$ -rich words.

**Proposition 21.** *Let  $\mathbf{u}$  be a ternary Arnoux–Rauzy word over the alphabet  $\mathcal{A} = \{0, 1, 2\}$  and  $\zeta$  be a binary projection over  $\mathcal{A}$ . If  $\mathbf{w} \in \{0, 1\}^{\mathbb{N}}$  is a preimage of  $\zeta(\mathbf{u})$  by  $S$ , i.e.,  $S(\mathbf{w}) = \zeta(\mathbf{u})$ , then  $\mathbf{w}$  is  $H$ -rich.*

*Proof.* We use the two well-known properties of Arnoux–Rauzy words which were already mentioned: any Arnoux–Rauzy word is uniformly recurrent and any Arnoux–Rauzy word



contains infinitely many palindromes centered at  $a$ , where  $a \in \mathcal{A} \cup \{\varepsilon\}$ . Therefore,  $\zeta(\mathbf{u})$  is uniformly recurrent and contains infinitely many palindromes centered at 1 and infinitely many palindromes centered at  $\varepsilon$ . Due to Proposition 19, the word  $\mathbf{w}$  is closed under all elements of  $H$ . According to Theorem 20,  $\mathbf{w}$  is  $H$ -rich.  $\square$

## 5 Comments and open questions

As already mentioned in the introduction, our computer experiments suggest that we may conjecture a more general statement than Theorem 3:

**Conjecture 22.** *Let  $\mathbf{u}$  be an episturmian word over  $\mathcal{A}$  and  $\zeta$  be a binary projection from  $\mathcal{A}$ . The word  $\zeta(\mathbf{u})$  is rich.*

Besides this conjecture, one may inquire about the class of binary words that is obtained by applying a binary projection on an episturmian word. Study of special factors, factor complexity and structure of return words may provide insight into this class. It would also be of interest to see if there is a relation to some known class of binary words.

## Acknowledgments

The author acknowledges financial support from the Czech Science Foundation grant GAČR 13-35273P. The computer experiments were performed using the open-source algebra system SAGE [33].

## References

- [1] P. ARNOUX AND G. RAUZY, *Représentation géométrique de suites de complexité  $2n + 1$* , Bull. Soc. Math. France, 119 (1991), pp. 199–215.
- [2] P. BALÁŽI, Z. MASÁKOVÁ, AND E. PELANTOVÁ, *Factor versus palindromic complexity of uniformly recurrent infinite words*, Theoret. Comput. Sci., 380 (2007), pp. 266–275.
- [3] L. BALKOVÁ, E. PELANTOVÁ, AND Š. STAROSTA, *Palindromes in infinite ternary words*, RAIRO-Theoret. Inf. Appl., 43 (2009), pp. 687–702.
- [4] —, *Sturmian jungle (or garden?) on multiliteral alphabets*, RAIRO-Theoret. Inf. Appl., 44 (2010), pp. 443–470.
- [5] —, *Infinite words with finite defect*, Adv. in Appl. Math., 47 (2011), pp. 562–574.
- [6] —, *Proof of the Brlek-Reutenauer conjecture*, Theoret. Comput. Sci., 475 (2013), pp. 120–125.

- [7] J. BERSTEL, *Sturmian and episturmian words: a survey of some recent results*, in CAI'07: Proceedings of the 2nd international conference on Algebraic informatics, vol. 4728 of Lecture Notes in Computer Science, Berlin, Heidelberg, 2007, Springer-Verlag, pp. 23–47.
- [8] A. BLONDIN MASSÉ, S. BRLEK, S. LABBÉ, AND L. VUILLON, *Palindromic complexity of codings of rotations*, Theoret. Comput. Sci., 412 (2011), pp. 6455–6463.
- [9] S. BRLEK, S. HAMEL, M. NIVAT, AND C. REUTENAUER, *On the palindromic complexity of infinite words*, Int. J. Found. Comput. Sci., 15 (2004), pp. 293–306.
- [10] M. BUCCI AND A. DE LUCA, *On a family of morphic images of Arnoux-Rauzy words*, in LATA '09: Proceedings of the 3rd International Conference on Language and Automata Theory and Applications, Springer-Verlag, 2009, pp. 259–266.
- [11] M. BUCCI, A. DE LUCA, AND A. DE LUCA, *Rich and periodic-like words*, in Proceedings of the 13th International Conference on Developments in Language Theory, Berlin, Heidelberg, 2009, Springer-Verlag, pp. 145–155.
- [12] M. BUCCI, A. DE LUCA, A. DE LUCA, AND L. Q. ZAMBONI, *On different generalizations of episturmian words*, Theoret. Comput. Sci., 393 (2008), pp. 23–36.
- [13] M. BUCCI, A. DE LUCA, A. GLEN, AND L. Q. ZAMBONI, *A connection between palindromic and factor complexity using return words*, Adv. in Appl. Math., 42 (2009), pp. 60–74.
- [14] —, *A new characteristic property of rich words*, Theoret. Comput. Sci., 410 (2009), pp. 2860–2863.
- [15] J. CASSAIGNE, *Complexity and special factors*, Bull. Belg. Math. Soc. Simon Stevin 4, 1 (1997), pp. 67–88.
- [16] D. DAMANIK AND L. Q. ZAMBONI, *Combinatorial properties of Arnoux-Rauzy subshifts and applications to Schrödinger operators*, Rev. Math. Phys., 15 (2003), pp. 745–763.
- [17] X. DROUBAY, J. JUSTIN, AND G. PIRILLO, *Episturmian words and some constructions of de Luca and Rauzy*, Theoret. Comput. Sci., 255 (2001), pp. 539–553.
- [18] F. DURAND, *A characterization of substitutive sequences using return words*, Discrete Math., 179 (1998), pp. 89–101.
- [19] A. GLEN AND J. JUSTIN, *Episturmian words: a survey*, Theoret. Inf. Appl., 43 (2009), pp. 403–442.
- [20] A. GLEN, J. JUSTIN, S. WIDMER, AND L. Q. ZAMBONI, *Palindromic richness*, European J. Combin., 30 (2009), pp. 510–531.

- [21] T. HARJU, J. VESTI, AND L. Q. ZAMBONI, *On a question of Hof, Knill and Simon on palindromic substitutive systems*, preprint available at <http://arxiv.org/abs/1311.0185>, (2013).
- [22] A. HOF, O. KNILL, AND B. SIMON, *Singular continuous spectrum for palindromic Schrödinger operators*, Comm. Math. Phys., 174 (1995), pp. 149–159.
- [23] J. JUSTIN, *Episturmian morphisms and a Galois theorem on continued fractions*, RAIRO-Theoret. Inf. Appl., 39 (2005), pp. 207–215.
- [24] J. JUSTIN AND G. PIRILLO, *Episturmian words and episturmian morphisms*, Theoret. Comput. Sci., 276 (2002), pp. 281–313.
- [25] J. JUSTIN AND L. VUILLON, *Return words in sturmian and episturmian words*, Theoret. Inf. Appl., 34 (2000), pp. 343 – 356.
- [26] S. LABBÉ AND E. PELANTOVÁ, *Palindromic sequences generated from marked morphisms*, preprint available at <http://arxiv.org/abs/1409.7510>, (2014).
- [27] S. LABBÉ, *A counterexample to a question of Hof, Knill and Simon*, preprint available at <http://arxiv.org/abs/1307.1589>, (2013).
- [28] M. MORSE AND G. A. HEDLUND, *Symbolic dynamics II - Sturmian trajectories*, Amer. J. Math., 62 (1940), pp. 1–42.
- [29] E. PELANTOVÁ AND Š. STAROSTA, *Languages invariant under more symmetries: overlapping factors versus palindromic richness*, Discrete Math., 313 (2013), pp. 2432–2445.
- [30] —, *Constructions of words rich in palindromes and pseudopalindromes*, preprint, (2014).
- [31] —, *Palindromic richness for languages invariant under more symmetries*, Theoret. Comput. Sci., 518 (2014), pp. 42–63.
- [32] G. RAUZY, *Suites à termes dans un alphabet fini*, Séminaire de Théorie des Nombres de Bordeaux, Anné 1982–1983 (1983).
- [33] W. STEIN ET AL., *Sage Mathematics Software (Version 6)*, The Sage Development Team, 2014. <http://www.sagemath.org>.
- [34] B. TAN, *Mirror substitutions and palindromic sequences*, Theoret. Comput. Sci., 389 (2007), pp. 118–124.